AMENDMENTS TO THE CLAIMS:

The claims are not further amended, and are presented below for the convenience of the Examiner.

Listing of Claims:

1. (Currently Amended) A method to process a document, comprising:

partitioning document text separated by spaces into a plurality of tokens based on the spaces;

identifying tokens to be ignored and not considered;

determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises:

examining syntax of the first token,

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying to the first token a plurality of regular expressions, rules, and a plurality of dictionaries comprised of a prefix dictionary, and a suffix dictionary to recognize chemical name fragments;

combining the first token with at least one of the adjacent tokens that are determined to be a chemical name fragment into a complete chemical name,

assigning the complete chemical name with one part of speech; and

storing in a memory the complete chemical name assigned with the one part of speech;

where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

2. (Original) A method as in claim 1, where the complete chemical name is assigned a noun phrase part of speech.

3-4. (Canceled)

5. (Original) A method as in claim 1, further comprising filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

6. (Original) A method as in claim 1, where chemical name fragments are further recognized by using common chemical word endings.

7. (Original) A method as in claim 1, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between chemical name fragments as a function of context.

8. (Original) A method as in claim 1, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

9. (Original) A method as in claim 8, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon.

10. (Original) A method as in claim 8, where the characters comprise at least one of upper case C, O, R, N and H.

11. (Original) A method as in claim 8, where the characters comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic.

12. (Original) A method as in claim 1, comprising an initial step of tokenizing the document to provide a sequence of tokens.

13. (Currently Amended) A system for processing a text document, comprising:

a first unit for partitioning document text separated by spaces into a plurality of tokens based on the spaces;

a second unit, operable for identifying tokens to be ignored and not considered;

a third unit, operable for determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises:

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying to the first token a plurality of regular expressions, rules and a plurality of dictionaries comprised of a prefix dictionary and a suffix dictionary to recognize chemical name fragments;

a fourth unit operable, to combine the first token with at least one of the adjacent tokens that are determined to be a chemical name fragment,

a fifth unit operable to assign the complete chemical name with one part of speech; and

a sixth unit operable for storing in a memory the complete chemical name assigned with one part of speech;

where the second unit is operable for identifying the tokens to be ignored by applying a negative dictionary to the plurality of tokens, and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

14. (Original) A system as in claim 13, where the complete chemical name is assigned a noun phrase part of speech.

15-16. (Canceled)

17. (Original) A system as in claim 13, where said second unit further comprises a sub-unit for filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

18. (Original) A system as in claim 13, where chemical name fragments are further recognized by using common chemical word endings.

19. (Original) A system as in claim 13, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between chemical name fragments as a function of context.

20. (Original) A system as in claim 13, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

21. (Original) A system as in claim 20, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon.

22. (Original) A system as in claim 20, where the characters comprise at least one of upper case C, O, R, N and H.

23. (Original) A system as in claim 20, where the characters comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic.

24. (Original) A system as in claim 13, further comprising a tokenizer for tokenizing the document to provide a sequence of tokens.

25. (Currently Amended) A computer program product embodied on a memory and executable to perform operations, comprising:

partitioning a document text ~~seperated~~ separated by spaces into a plurality of tokens based on the spaces;

identifying tokens to be ignored and not considered;

determining that a first token considered of the plurality of tokens comprises an organic chemical name fragment, wherein determining comprises:

examining syntax of the first token,

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying a plurality of regular expressions, rules, and a plurality of dictionaries comprising a prefix dictionary, and a suffix dictionary to recognize organic chemical name fragments;

combining the first token with at least one of the adjacent tokens that are determined to be an organic chemical name fragment into a complete organic chemical name;

assigning the complete organic chemical name with one part of speech; and

storing in a memory the complete organic chemical name with the one part of speech;

where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

26. (Original) A computer program product as in claim 25, where the complete organic chemical name is assigned a noun phrase part of speech.

27-28. (Canceled)

29. (Original) A computer program product as in claim 25, further comprising instructions for filtering recognized organic chemical name fragments using a list of stop words to eliminate erroneous fragments.

30. (Original) A computer program product as in claim 25, where chemical name fragments are further recognized by using common chemical word endings.

31. (Original) A computer program product as in claim 25, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between organic chemical name fragments as a function of context.

32. (Original) A computer program product as in claim 25, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

33. (Original) A computer program product as in claim 32, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon, where the characters comprise at least one of upper case C, O, R, N and H, and further comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic.

34. (Original) A computer program product as in claim 25, where said instructions for assigning operate on a sequence of tokens derived from document text.

35. (Currently Amended) A system comprising a plurality of computers at least two of which are coupled together through a data communications network, said system comprising

a first unit for partitioning document text ~~seperated~~ separated by spaces into a plurality of tokens based on the spaces;

a second unit, operable for identifying tokens to be ignored and not considered;

a third unit, operable for determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises:

examining syntax of the first token,

examining context of the first token with respect to at least one adjacent token of the plurality of tokens, and

taking into account the syntax and the context, applying a plurality of regular expressions, rules, and a plurality of dictionaries comprised of a prefix dictionary and a syntax dictionary to recognize chemical name fragments;

a fourth unit, operable to combine the first token with at least one of the adjacent tokens that are determined to be a chemical name fragment into a complete chemical name;

a fifth unit, operable to assign the complete chemical name with one part of speech; and

a sixth unit, operable for storing in a memory information the complete chemical name with the one part of speech;

where the second unit is operable for identifying the tokens to be ignored by applying a negative dictionary to the plurality of tokens, and wherein the plurality of dictionaries consists of the prefix dictionary, the suffix dictionary, and the negative dictionary.

36. (Original) A system as in claim 35, where the complete chemical name is assigned a noun phrase part of speech.

37. (Original) A system as in claim 35, where a user of the system accesses the system through a data communications network.

38-39. (Canceled)